

REVIEW

A Review of Strategies for Designing, Administering, and Using Student Ratings of Instruction

Melissa S. Medina, EdD,^a W. Thomas Smith, PharmD, JD,^b Srikanth Kolluru, PhD,^c
Elizabeth A. Sheaffer, PhD, MBA,^d Margarita DiVall, PharmD, MEd^{e,f}

^a University of Oklahoma Health Sciences Center, College of Pharmacy, Oklahoma City, Oklahoma

^b Manchester University College of Pharmacy, Natural and Health Sciences, North Manchester, Indiana

^c Keck Graduate Institute School of Pharmacy, Claremont, California

^d Samford University McWhorter School of Pharmacy, Homewood, Alabama

^e Northeastern University School of Pharmacy, Boston, Massachusetts

^f Editorial Board Member, *American Journal of Pharmaceutical Education*, Arlington, Virginia

Submitted May 17, 2018; accepted January 30, 2019; published June 2019.

Objective. To review and recommend strategies for utilizing student ratings of instruction (course and instructor) including considerations regarding design, administration, and use and interpretation of results.

Findings. Improving course delivery and pedagogy using student ratings of instruction requires programs to design evaluation instruments that are aligned with the following good, scholarly teaching criteria: offer 10-20 rating scale questions and at least one written response question, ensure that students understand what the questions are asking, use a standardized form for evaluating all faculty members, allow for additional tailored questions to be added to the form, and employ a four- or five-point rating scale with a “not applicable” option. When administering evaluations, programs should limit the number of faculty members evaluated to those teaching greater than or equal to five clock hours of lecture or schedule evaluations based on academic rank; use an online course evaluation tool; randomly select students to participate; offer the evaluation at the end of the term (and/or midpoint for team taught classes); offer the evaluation during scheduled class time; and allow for voluntary, anonymous student participation. Finally, programs should create an assessment plan that outlines the results’ release timeline, a list of who will receive result summaries, and how the results will be used. Programs should also encourage faculty reflection, offer mentoring in results interpretation, coach faculty members to summarize and quantify comments and longitudinally track results using tables, and create an accountability action plan to address deficiencies.

Summary. In order to better ensure that student ratings of instruction are used to improve teaching, colleges and schools should adopt intentional design, structured administration processes, and transparent reporting of results.

Keywords: course evaluation, assessment, evaluation, faculty

INTRODUCTION

Student ratings of instruction, also known as “course evaluations,” are the most common way students provide feedback about faculty teaching and course design and delivery, regardless of discipline, program, degree awarded, or institution type.¹ One primary reason that

student ratings of instruction are administered is for continuous quality improvement of courses and faculty teaching. Gathering student perceptions of teaching and course delivery is important because students are the direct recipients of the instruction and can offer important insights regarding the learning and assessment process and how teaching can be improved.³ Institutions also administer student evaluations of instruction to meet specific or implied regional and/or professional accreditation requirements.⁴⁻¹⁰

Despite being one of the most common and efficient ways to gather broad student feedback, student ratings of

Corresponding Author: Melissa S. Medina, Department of Pharmacy: Clinical and Administrative Sciences, The University of Oklahoma College of Pharmacy, P.O. Box 26901, 1110 N. Stonewall, CPB 125, Oklahoma City, OK 73190. Tel: 405-271-6484. Fax: 405-271-3830. Email: melissa-medina@ouhsc.edu

instruction are one of the most scrutinized and debated topics in higher education.³ There are over 3000 publications dedicated to the topic, yet misperceptions, questions, and concerns about student ratings of instruction design, administration, and use of results still persist.² Three areas in particular, design, administration, and use of results, remain contentious issues in higher education and health professions programs because results are often used for annual performance reviews and promotion and tenure decisions.¹¹ The continued debate regarding the use of student ratings of instruction warrants an updated review of strategies related to these three central areas (design, administration, and results use). The last comprehensive review of this topic as it relates to pharmacy education was published in 2009.⁴ More recent pharmacy publications in this area have focused on narrow research areas such as factors influencing student completion of evaluations.¹²

METHODS

To gather information on student ratings of instruction strategies, a literature review was conducted using Ovid MEDLINE, PubMed, and PsycINFO. Search terms used included: “course evaluations” and “student ratings of instruction” in an effort to focus the review. The search produced 795 articles for “course evaluations” and 104 articles for “student ratings of instruction.” Articles were included if they were published in the last 10 years (2008-2018), written in English, described the process of administering student ratings of instruction, and focused on health professions education. Making this distinction was important as there are some differences between health professions curricula and undergraduate curricula in course selection choice, teaching methods, number of faculty members teaching in a course, and testing format, which all may influence the student ratings.¹⁴ A database of eligible studies was compiled and sorted by date (greater or less than 10 years old) and topic area (forms, administration, response rate, and results use). Additional studies were included that were older than 10 years because of their specific emphasis on pharmacy education and longstanding teaching and assessment theory.^{1,4} Articles were reviewed and selected for inclusion in the paper based on agreement among the five coauthors. Articles were excluded if they focused on using student ratings of instruction for research and assessment purposes (eg, gathering feedback about course satisfaction related to a research study). Articles were also excluded if they focused on other sources of data about teaching quality and effectiveness, such as peer review and self-assessment, because these topics warrant a separate review. Although triangulating teaching feedback from

students, self, and peers is vital for continuous quality improvement,¹³ this review concentrates on student ratings because of the ongoing concerns over the use of student ratings of instruction design, administration, and results for faculty evaluation and course revisions.

The information provided is organized into three main sections and includes challenges in and strategies for designing, delivering, and using student ratings of instruction. Topics were determined based on themes found in the literature and confirmed by the authors. The strategies reviewed in this paper are intended for higher education and health professions programs, though each institution must consider what works best in its specific setting. There are varying challenges and levels of control over the design/format, delivery, and use of the rating instruments and systems depending on university size, resources, and type of program. These challenges as well as recommendations for each area are described.

Designing Student Ratings of Instruction

Question creation. Determining what questions to ask is the first step in conducting student ratings of instruction because the right questions can provide data that help improve teaching and learning.³ Addressing this step is challenging because programs must define teaching effectiveness or outline the criteria for good teaching¹⁵; however, definitions of good teaching vary and the completeness of those definitions influence what questions are asked.¹⁶ Programs could consider using the concept of scholarly teaching to outline criteria for good teaching.^{17,18} Scholarly teaching is defined as the use of effective teaching methods which lead to student learning.¹⁹ Scholarly teachers use evidence-based, systematic teaching methods and possess three types of knowledge: content knowledge (knowledge of one’s discipline), pedagogical knowledge (how to teach within one’s discipline and what makes the learning of specific topics easy or difficult), and curricular knowledge (understanding how one’s course topics relate to and affect other courses within the program).¹⁷⁻²⁰ There are six standards for evaluating scholarly work and scholarly teaching: clear goals, adequate preparation, appropriate methods, significant results, effective presentation, and reflective critique.²⁰

Recommendations for question creation. The foundation of a course evaluation is the questions that are asked, although there is great variability in what questions are asked, as no set standard exists. This lack of consensus in academia is most likely related to the lack of agreement with how to define and measure teaching effectiveness.³ The literature offers little guidance on creating questions. DeCourcy reports that quantitative student ratings are the most commonly used method to evaluate student

perceptions, and that using six to nine criteria is generally reliable.¹⁶ The six standards of scholarly teaching is one example of a framework that can be used to create the question categories and quantitative questions for the student rating of instruction form because that is a framework recommended to evaluate teaching effectiveness in the pharmacy literature.¹⁷⁻²¹ For example, evaluation questions can be asked for each of the standards (Table 1).^{17,21} Other examples of question categories reported in the literature are course organization and planning, communication skills, teacher student interaction, course workload, course grading or testing, and students self-rated learning; however, these categories could be mapped onto the six standards described above.² In the literature search, the authors found the Course Experience Questionnaire, which is a form that demonstrated validity and reliability in medicine and measures five areas: quality of teaching, clarity of goals and standards, nature of assessment, level of workload, and development of generic skills.²³ However, the questions for this form were not readily available and did not easily map onto the six standards of scholarly teaching. Overall, a review of the literature revealed that there is variability in the questions and categories that are asked on student ratings of instruction forms, which is most likely related to the lack of agreement in teaching effectiveness definitions.³ Ryan and Harrison found that the two most important criteria students used in judging teaching effectiveness are perceptions of examination fairness and amount of content learned, which could also be mapped onto the scholarly teaching category of significant results.²⁴ Therefore, pro-

grams should consider using questions related to these two areas. To assist students with being able to assess their learning, it may be best to ask them to evaluate how much progress they have made on learning objectives for the course (eg, no apparent progress to exceptional progress), but this approach may require programs to customize student rating forms.²⁵ Another way of asking students about learning in the course is to ask their perceptions of course organization around learning objectives.²⁶ In addition, the literature also recommends that summary rating items, such as the overall quality of the teaching and course, as well as open-ended summary items, be included on the form because these questions complement student ratings and provide detailed information about faculty strengths and weaknesses.²³

Question clarity. One challenge when developing the student ratings of instruction questions is ensuring that respondents understand what they are being asked to evaluate. Few studies have evaluated students' perceptions of the questions used in the evaluation process. One study that included 330 second-semester students from the University of Teknologi Mara found that only 42.4% of students felt evaluation questions were clear in a given course evaluation.²⁷ It is difficult for educators to acquire meaningful results from evaluations if educational jargon or vague language interferes with students' understanding of question intent. It is also important to consider whether students are able to judge a specific characteristic of instruction. For example, students may not be able to judge the instructor's degree of expertise in a particular topic, but they are able to comment on an instructor's

Table 1. Course and Instructor Evaluation Questions Related to the Six Standards of Scholarly Teaching

Six Standards of Scholarly Teaching	Sample Questions for Course Evaluations	Sample Questions for Instructor Evaluations
Clear goals	Created measurable and quantified course objectives	Created measurable and quantified lecture objectives ¹⁷
Adequate preparation	Stated the relevance of course to professional career	Activated students' prior knowledge ¹⁷
Appropriate teaching methods	Used a variety of teaching methods Structured and organized content Reinforced course concepts	Used active learning ¹⁷ Gave multiple examples and/or pointed out practical applications ¹⁷
Significant results	Aligned test questions with objectives Used grading tools such as rubrics	Used frequent assessments to assess knowledge and skills ¹⁷
Effective presentation	Offered students timely feedback "My curiosity was stimulated" ²² "I am more interested in the subject matter" ²² "I am learning to think more clearly" ²²	Used handouts and visual aids Engaged students ¹⁷ Asked and answered questions ¹⁷
Reflective critique	Encouraged reflection on learning throughout the course	Encouraged reflection on learning from the lecture ¹⁷

enthusiasm about the topic and the ability to provide practical examples of concept application.

Recommendations for question clarity. One practical strategy for programs to use to ensure or improve the clarity and intent of the evaluation questions is to obtain feedback from randomly selected groups of students. This process can help clarify questions and positively affect the ability of faculty/administrators to interpret results of the evaluation at various levels of review.

Number of questions. After a program defines teaching effectiveness and a related framework for the questions, programs must decide how many questions to ask on the rating instrument. Determining the number of questions is important because too few questions may not give faculty members enough feedback to improve their course or teaching and too many questions may create survey fatigue and decrease response rates. Guidelines regarding the optimal number of questions to include on evaluation instruments are lacking, which makes it difficult to offer a definitive recommendation of an exact number or range. One small focus group study evaluated the perceptions of 17 medical students regarding the length of evaluations and found that students preferred being asked no more than 15 questions.²⁸ Other reports only document the number of questions researchers used in their studies or reviews. For example, Anderson and colleagues¹ recommended 21 questions covering four areas: the course (eight items), instructor (six items), learning outcomes (five items), and summary (two items). Meyer and colleagues developed and validated a 30-item criterion-based instrument covering organization and structure (six items), assessment and feedback (six items), personal interactions (four items), and academic rigor (nine items).²⁶ The 2012 American Association of Colleges of Pharmacy's Academic Affairs Committee recommended evaluating teaching excellence using the scholarly teaching framework based on Glassick and colleagues' six standards of scholarly teaching.^{17,18,21} Using the six standards of scholarly teaching framework, this report recommended 10 questions that programs can use for students to answer about the course: clear goals (one question), adequate preparation (one question), appropriate teaching methods (three questions), significant results (two questions), effective presentation (one question), and one overall summary rating and one written course comment (see Table 1 for an example of how to create questions using one framework).^{17,18} Using the same framework (Table 1), students could address 13 additional questions pertaining to the instructor, including clear goals (one question), adequate preparation (two questions), appropriate teaching methods (one question), significant results (two questions), effective presentation

(four questions), reflective critique (one question), one overall summary rating, and one written comment (please provide constructive comments about the instructor).^{17,18,21} Overall, given the multifaceted nature of teaching and course delivery, 15-30 questions appears to be a common length for student ratings of instruction. In programs that use team-taught courses, separating the evaluation of the course from the evaluation of the instructors may reduce the total number of questions because students would only be repeating the instructor questions. Programs should consider stating in the instructions how many total questions will be presented to them or giving students a completion progress indicator for electronic evaluations so that they are able to estimate time needed for completion.

Standardized questions. Once questions are created, programs need to determine if all instructors and courses should use the same queries even if the courses are delivered differently (eg, laboratory, experiential, lecture-based) and for all faculty members from different departments in the same college or school or different disciplines across the university.³ Using the same questions allows for comparisons. However, ratings vary by discipline, so if comparisons are made across courses, disciplines, or departments, scores should be reported with standard deviations from the respective area.³ One disadvantage of this approach is that using the same questions may not fit the needs of every course and instructor, which may interfere with the usefulness of these results to improve teaching. This can be particularly problematic in health professions education, where the format of courses can vary from large didactic courses, to small-group laboratory and seminar courses, to experiential courses. For example, an experiential course evaluation in which the majority of questions were related to clinical practice may be of limited utility for an elective rotation in research or academia. Additionally, when courses are delivered using drastically different methodologies (eg, lecture vs flipped classroom), a standard student rating form may not be of value to all instructors.

Recommendations for standardized questions. Programs should be flexible with what questions are asked, otherwise the system can be unfair for all.²⁹ One way to offer flexibility is to have standardized core evaluation questions and allow instructors to include a limited number of additional questions at the end of the survey instrument. These additional questions may be custom or selected from a defined list and pertain to course objectives or type of course, educational setting, and/or instructional methodology.^{22,30,31} Limiting the number of these additional questions is important in order to mitigate survey fatigue. This flexibility is useful because it offers

faculty members pertinent feedback they can use to improve their course or teaching.

Programs might consider the use of a mid-course or midpoint evaluation for faculty members to gather formative feedback about their course(s). The advantage of this process is that it allows faculty members to ask specific questions relevant to their course and gather timely, formative feedback about their teaching or course that could result in modifications for that term benefitting the same students enrolled in the course. Faculty members could create their own questions for this midpoint evaluation rather than using a standardized set of college or school questions as the purpose of the data gathered would be for personal reflection and use and possible course modification. These results would not be intended for comparison with those of other faculty members. The disadvantage of midterm evaluations is that they would present an additional workload, although some programs have student class representatives and/or class officers lead and manage this initiative. If students manage the process, training in administering evaluations would need to be in place to ensure anonymity of the results. Mid-course evaluations also may not be practical for block-style courses in which the content is delivered in a much shorter timespan.

Rating scales. The Likert scale is the most commonly used psychometric scale for student ratings of instruction (eg, 1=strongly disagree, disagree, neutral, agree to 5=strongly agree; or 1=poor, fair, good, very good, 5=excellent), although the appropriateness of calculating means from ordinal data has been debated.^{4,11} There is disagreement in the literature about which rating scale to use. Some studies suggest using either a five-point or seven-point scale because using less than five points does not discriminate well and more than seven points does not offer additional value.³² In contrast, Peeters recommends using a four-point rating scale because there may not be five, six, or seven distinct categories.³³ Additional debates also focus on including a “neutral” option and/or a “not applicable” option because some students may be undecided, have no opinion, or feel the outcome was truly average. Having a “not applicable” option also may be beneficial for laboratory or experiential courses where items on the standardized rating form may not be relevant (eg, the instructor’s slides and/or handouts facilitated learning).

Recommendations for rating scales. The number of rating scale points is a topic that is highly debated. A four- or five-point scales may offer students clearer distinctions compared to a seven-point scale. This five-point scale can help faculty members better understand distinctions among student perceptions. If programs observe that a majority of students primarily select the neutral/average

option, then the program should consider using a four-point scale that does not include a neutral/average option. Program administrators should also include the “not applicable” or “unknown” option in standardized scales if questions may not be relevant to all courses.

Administering Student Ratings of Instruction

Number of faculty to evaluate in a course. In comparison to undergraduate courses, health professions education frequently employs team-taught courses to leverage specific expertise of faculty members. For team-taught courses, inquiries arise about which faculty members should be evaluated. The more faculty members that are included in the evaluation, the more time it will take students to complete the evaluation, which could negatively affect response rate.¹² For example, students can be asked to complete anywhere from 12 to 20 course evaluations and evaluate 24 to 60 instructors annually depending on the structure of the curriculum.¹¹

Recommendations for number of faculty members to evaluate. When deciding how to limit the number of faculty members evaluated in a team-taught course, the authors recommend only evaluating faculty members who teach at least a certain number of hours or percent of time in the course, for example, at least four to five hours or 20% of a course.⁴ A second option that can be added to the above recommendation is to consider the rank of the faculty members in the course: new or non-promoted faculty members receive more frequent evaluation (eg, every semester or yearly) because the results may be needed in making promotion or contract renewal decisions, while promoted and/or tenured faculty members might receive an evaluation on a schedule (eg, every three years).⁴

Technology used. Course evaluations can be administered using various platforms. Web-based (online), electronic survey tools are now the most common way to create and deliver course evaluations.⁴ They are generally preferred over paper-based evaluations because electronic tools save time, paper, and personnel resources, and allow for quick dissemination of feedback to faculty members after the evaluation closes.⁴ The transition from a paper-based to an electronic evaluation system is not simple. Students may prefer online evaluations because of their flexibility and lack of time limits, whereas faculty members may prefer paper-based evaluations because of concerns over response rates and representativeness.¹⁴ Paper evaluations often yield a 70%-80% response rate, whereas response rates for electronic evaluations are often significantly lower.²³ Some faculty members argue for a return to the use of paper evaluations to increase evaluation response rates so that the results are more representative.^{23,34} As a result of this debate, one study that

included 81 instructors and 247 course sections from undergraduate, graduate, and professional degree programs at one university randomized 4550 students to receive either an electronic (2280) or paper-based evaluation (2270). The study found lower response rates for the electronic version of the evaluation, however, scoring patterns and mean scores were similar for both methods.³⁴ Another study from six departments in the school of business at Loyola University compared 4424 paper and electronic evaluations. Consistent with the previous study, there was a lower response rate for online evaluations but no significant differences in instructor and course ratings between the two evaluation methods. Also, students who completed the online version provided more and lengthier comments.³⁵ A third, smaller study found no statistical difference between mean scores on course evaluations delivered online even though there was a higher response rate for paper evaluations.³⁴ What these study results consistently suggest is that online evaluations are a suitable alternative for conducting evaluations as scoring patterns are similar even if response rates are lower than with paper evaluations. Options for conducting electronic evaluations include using free survey tools such as Google Forms (Google LLC, Mountain View, CA) and SurveyMonkey (SurveyMonkey, San Mateo, CA), which are limited to 100 participant responses with a free license. Electronic survey instruments can also be created and distributed by learning management systems such as Blackboard (Blackboard Inc, Washington, DC) or Desire to Learn (D2L Ltd, Towson, MD), that may already be financially supported by the University. There are also options to purchase licensed software such as CourseEval (Invoke Solutions, Waltham, MA), E*Value (MedHub, Minneapolis, MN), and Qualtrics (Qualtrics LLC, Provo, UT).

Recommendations for technology. The authors recommend using online survey systems rather than paper and pencil for students to rate courses and instructors because of their cost-effectiveness, environmental friendliness, quick distribution and results dissemination, and the availability of data analytics software.³⁴ Although paper evaluations may yield a higher response rate, research studies have demonstrated the equivalence in results between the two evaluation methods. Colleges desiring to increase survey response rates should educate faculty about the results equivalence and assure students of the anonymity of the results. Another option colleges should consider to increase response rates is the use of a random sample of students from within a given cohort to complete the evaluation versus administering the survey instrument. Random samples are likely to yield the same reliability when compared to surveying the entire popu-

lation of students.²³ While each online survey tool described above offers different features and benefits, the authors recommend using a tool with specific course evaluation features, such as CourseEval or E-Value, because these tools allow for anonymity of responses and ease of distribution, analyzing, reporting, collating, and archiving evaluation results in a timely manner. In contrast, while using Qualtrics or a learning management tool to administer student surveys may be efficient and allow anonymity, these tools do not allow for easy collating and reporting of results across multiple years or courses.

When to administer evaluations. The most common option for administering evaluations of semester- or quarter-long courses to students is at the end of the term, specifically, during the last two weeks of class prior to the final examination.^{12,32} Administering evaluations at this time allows students to form the most complete perceptions about the course and what was learned. Also, students may have fewer competing deadlines at this time.³² The disadvantage is that some students prefer to complete evaluations after the course's final examination because at that point they will know their final course grade and/or have more time to complete the evaluation. Offering the evaluation after course grades are calculated and/or final examinations are completed may confound results as students may use their course ratings as a way to reward or penalize faculty members based on the grades they earned.²³ However, research has shown that the point during the school term at which the evaluation is offered is not correlated to student ratings.² For example, in one study, no difference in ratings were seen when evaluations were offered any time of day and at any time during the second half of the course, the middle versus end of the term, the last week of class versus the first week of the next term, or the last day of class versus the day of the final examination. However, administration on the day of the final is discouraged because the evaluation may only reflect feeling of the final examination.²

In addition to determining when to open evaluations, programs must also consider when to close evaluations. There is variability in opinions of how long the evaluation window should remain open. Some programs may allow students to complete the evaluations over a two-week window prior to examinations, while others may close the evaluations at the end of the final week or the week after final examinations end.

Recommendations for when to administer evaluations. There is a lack of literature evaluating the impact of a given time window (how long to leave the evaluation open) on course evaluation results, it is recommended to offer the evaluations at the end of the semester.³⁶ One exception is for team taught courses: programs may want

to select an earlier timeframe, such as the midpoint of the course, to capture feedback about faculty teaching in the first half of the term so that students do not forget or confuse the instructors when completing evaluations. The date of the in-class evaluation should be noted on the course schedule at the beginning of the semester. Also, students should be offered class time (at least 20 minutes) to complete course evaluations so that they are spread out among courses to reduce student survey fatigue.³⁶

Faculty members should summarize previous evaluation results and subsequent changes made to demonstrate to students that their feedback is valued as this may encourage students to complete evaluations.³⁶ For example, a faculty member might explain that previous students suggested that he/she hold a review session prior to the final examination and since he/she has done so, students' examination grades have improved. To complement this in-class explanation, a description of how the survey results will be used, ie, for formative purposes, such as for teaching or course delivery improvement, and/or for summative purposes, such as instructor performance evaluation, could be included in program materials and course syllabi.³⁵ Faculty members could also offer examples of what a useful student comment might be, eg, "It would have helped me focus my studying and learning better if you would have offered more specific and quantified lecture objectives." In contrast, the faculty member could explain to students that comments that are vague or judgmental are not helpful in understanding what exactly needs to change, eg, "Your class was so boring that I slept through most of the lectures."

Having a neutral party make the evaluation available to the students and ensuring that the faculty member leaves the room if students are given time to complete the evaluation during class are also important. Research has shown that ratings are inflated when the faculty member is present in the room while students are completing an evaluation.² Another alternative is to only allow the evaluation to be completed during class time on a designated day at the end of the semester in order to mimic traditional paper and pencil administrative procedures. This would negate the need for a time window. If a course is team taught then a midpoint day can be used to collect evaluations for faculty teaching earlier in the course.

Required vs voluntary evaluations. Ensuring class representativeness (ie, a high response rate) is a desired outcome when administering and collecting student ratings of a course or instructor.³ Some programs believe that the way to improve response rates for student evaluations is to require mandatory completion, although this has not been well studied.²² Some programs use a stick approach (eg, withholding grades) while others use a car-

rot approach (eg, providing individual or class-wide incentives) to ensure completion and high response rates. The benefit of requiring students to complete a course or instructor evaluation is an increased likelihood of capturing all levels of student satisfaction, whereas with voluntary evaluations significantly more students who are extremely satisfied or dissatisfied may respond than students with neutral feelings. The main disadvantage of requiring students to complete course evaluations is that the process can be interpreted as coercion. As stated in Federal Regulations (45 CFR 46.116) and Institutional Review Board policies, students are a protected/vulnerable population and should be given the opportunity to consider whether to participate in research and minimize the possibility of coercion or undue influence.^{39,40} Therefore, penalizing students for refusing to participate in the course evaluation process could be viewed as coercion or undue influence.⁴⁰ Second, as mentioned in the technology section, low response rates may still yield meaningful data for improving teaching, especially if similar themes are present in the results across multiple courses and years.³

Recommendations for requiring evaluations. Although requiring students to complete evaluations may increase response rates, this method might be used with caution because requiring students to complete an evaluation could be interpreted as coercion. While punishments such as having to meet with the dean or withholding items such as grades may be obvious forms of coercion, offering rewards and incentives such as extra points for completing evaluations could be perceived as coercion as well even though some participant incentives are allowed in research. Instead, programs may want to identify a random but representative sample of students (eg, at least two-thirds of the class) to complete the evaluations.²² Based on published reports, offering evaluations to the entire class versus a randomly selected sample may not provide better reliability.²³ If response rates remain low, programs may want to increase buy-in from faculty members and students about course evaluations and gather suggestions about how to increase response rates so the process can become part of the culture.

Interpreting and Using Student Ratings of Instruction Results

Results availability and access. Once students complete course and instructor evaluations, procedures should be in place that indicate who should have access to the results.³⁷ The plan should outline when faculty members will receive evaluation results, eg, within one month after the evaluation window closes and after final course grades are submitted. The plan should also outline who in

addition to the faculty members involved with the course receives the results. Results are commonly distributed to department chairs for use in faculty members' annual reviews, but questions exist as to whether deans (including associate and assistant deans), curriculum and/or assessment committees, and peer mentors should also receive the results. Sharing results with students or the general public is uncommon unless perhaps teaching is perceived as a program strength at a particular school and results could be used for recruitment.^{4,38} Colleges should outline and distribute a transparent plan to all of its stakeholders in the assessment plan in order to close the assessment loop and so faculty members can use the results to fulfill the goal of the evaluations, which is to improve their courses and teaching.

Gender and race bias in results. All people accessing student evaluations of teaching results should be aware of controversies surrounding student ratings of instruction associated with non-modifiable factors such as race and gender and consider these when interpreting evaluation results. One literature review found that student ratings of instruction are not affected by the teacher's age, gender, race, or personal characteristics.² However, other studies indicated differences in ratings that were explained by sex and gender.⁴¹ For example, through content analysis of student evaluation comments, one study found that students used significantly different language in evaluating female vs male professors on intelligence/competence, personality, and appearance.⁴² Women were more likely to be referred to as a "teacher" and called "Mrs" compared to men who were more likely to be referred to as a "professor" and addressed as a "Dr," which suggests the students perceived men to be of higher rank and competence.⁴² In addition, students commented on a woman's appearance more often than a man's.⁴² Other studies report similar findings of bias against women and minorities.⁴³⁻⁴⁵ Further research about bias in teaching and student evaluations of teaching is needed. In the meantime, mentors, department chairs, administrators, and all faculty members serving on merit and promotion and tenure committees should be aware of the potential for bias when teaching evaluation results are used for formative and/or summative reasons. Pharmacy schools should also work to reduce bias by increasing students' and faculty members' awareness of bias in the classroom and on course evaluations, using inclusive language on the evaluation forms (ie, s/he), and acknowledging how bias may be present during faculty teaching reviews.⁴³

Course coordinator results access. In programs that have team-taught courses, the course coordinator should have access to the results of evaluations of faculty members who teach the course in order to improve teaching

within a course and overall course delivery. This access allows the course coordinator to evaluate any themes present in the evaluations across all the faculty members involved in teaching the course. For example, the coordinator may discover that multiple faculty members are unintentionally teaching the same content or delivering conflicting information about the same content. This discovery will allow the coordinator to promote teaching improvement and discuss strategies with the faculty members for eliminating redundancies, reconciling competing facts, and using intentional repetition and connection of important course concepts. Course coordinators should receive training with regards to the use of evaluation results as a means to provide feedback to faculty members with the goal of improving course delivery.

Mentor and administrator access to evaluation results. Receiving consultation or mentoring about evaluation results improves teaching.³ Peers or educational consultants can offer this mentoring, but they should receive training in how to provide effective course evaluation feedback. Mentors can help faculty members interpret results, put negative results into perspective, and develop realistic goals for improvement. Department chairs may request the raw data or a summary of results during an annual review. The results may be used for formative reasons such as to help with results interpretation, offer mentoring, or help identify faculty development programs to facilitate teaching improvement, similar to peer consultants described above. The results may also be used for summative purposes such as to determine promotion and/or tenure progress, merit raises, or teaching loads, but the student ratings of instruction should be considered along with other factors and not be the sole indicator for such determinations. Faculty members may feel less comfortable receiving formative mentoring and guidance from the department chair because he or she uses the results for summative reasons.

The administrator responsible for teaching, assessment, academic affairs, curriculum, and/or the assessment committee can use course and instructor evaluation results to track assessment and/or strategic plan initiatives, or to identify college-wide faculty development programs to improve teaching. Whether the committee(s) should have access to the raw data or the faculty member being evaluated should provide committee members with a summary of the evaluation is debatable as requiring committee members to review all student evaluation results and comments regarding the faculty member would be too time consuming. Whenever interested stakeholders do not have access to raw student evaluations, faculty members can be asked to

share their longitudinal tables and/or a summary of student comments regarding their strengths and areas for improvement.

Faculty results interpretation. Reviewing results is not an entirely intuitive process, and individual numerical scores or poorly worded student comments may distract faculty members. Therefore, faculty members should receive guidance on interpreting results, engaging in self-reflection, and tracking/comparing their results over time in order to improve the course delivery and teaching.³¹ For example, a faculty member should be asking himself/herself questions like: “Are the student results consistent with my experience in the course?” “How might I teach the course differently next time to improve student learning?”³ While some programs may want benchmarking data where faculty members compare their results (eg, mean scores and standard deviations) to those of other faculty members in their department, discipline, college/school, or university, these comparisons can be skewed because ratings can vary by discipline.³ Therefore, faculty members should benchmark against themselves and evaluate how they are performing over time. Reviewing results from student ratings of instruction is intimidating but comparing their data over time can help faculty members objectify their results and focus on using the results to improve teaching. One strategy that the authors recommend as an example is to use a written comments summary tool to categorize and quantify the themes in their results (Appendix 1). The authors further suggest that faculty members can also interpret their numeric results by summarizing them in a longitudinal table (Appendix 2) and looking for patterns of strengths and areas of improvement. Faculty members can use these longitudinal summary data to set future teaching goals and track improvements for personal use, reflection in teaching philosophies, and/or documentation in a teaching portfolio, department annual reports/review, and/or promotion/tenure.¹³

Faculty action plans. One formative strategy to improve teaching and/or course delivery is the use of course and instructor evaluation summaries and action plans. A study by Fleming and colleagues evaluated whether an intensive course review protocol improved future delivery of a given course.³⁷ In this study, after course coordinators reviewed their course evaluation results, they sent a written summary of the results within one month to their program evaluation subcommittee, which is similar to a curriculum or assessment committee. A course underwent an intensive review if the results did not meet established standards, including an overall mean rating below 3.5 (on a scale of 1 = low to 5 = high); a greater than 0.5-point drop in the mean course rating over one year; or a committee-

identified critical issue such as poor examination results. For the intensive course review, the course coordinator created an action plan to address the deficiency and resolve the identified problem(s). Student liaisons offered input about the plan and changes were tracked over time until improvements were documented. The authors identified three benefits of this process. First, it characterized a systematic process for identifying courses with negative student comments or rating using predetermined benchmarks in order to promote course or teaching improvement. A second benefit was that student representatives were included in the process, which helped students understand the impact of completing their evaluations and allowed them to offer suggestions for improvement. Finally, this process emphasized transparency and course coordinator accountability to improve teaching by requiring the development of a formal action plan. This study found that the intensive course review had a positive impact on course delivery as well as on future course ratings. However, the benefits were limited by the overall course and instructor evaluation response rates. Low student participation in the student ratings of instruction threatened to negatively impact the intensive course review process if the results did not truly reflect the majority of students’ perceptions.³⁷ Overall, the summaries of faculty member and course evaluations are a useful formative process for promoting improvement in teaching and instruction, but the process relies on faculty members creating the course and instructor summaries, which could be a rate-limiting step. Additionally, programs need to consider if and/or what consequences faculty members should face if they receive poor teaching evaluations over a period of time and do not attempt to make any teaching improvements.

CONCLUSION

Whether students should evaluate instruction remains a highly debated topic even as new strategies for creating, conducting, and using the results of evaluations continue to evolve. While the literature in higher education offers some guidance for how to use evaluation results to improve teaching and courses, the nature of health professions education offers some unique challenges that require additional consideration. To achieve the ultimate goal of student ratings of instruction, ie, to improve teaching and courses, programs need to first define the criteria for good teaching (eg, scholarly teaching) before they design an instrument to evaluate it. In the instrument design, they should create questions aligned with the criteria for good teaching (eg, use the six standards of scholarly teaching to create and categorize questions); limit the number of questions to 10 to 20 to help mitigate student survey fatigue; actively ensure that students understand

what the questions are asking; use the same questions for all faculty members, but allow instructors to include additional tailored questions; and employ a four- or five-point rating scale with a not applicable option. Once the evaluation form is designed, programs should limit the number of faculty members who are evaluated to those teaching approximately five hours in the course and/or schedule faculty members for evaluations according to academic rank (junior faculty=every semester; senior faculty=every 3 years); administer the evaluation using an online course evaluation tool; randomly select students to complete the evaluation; open the evaluation at the end of the semester/term and/or at the midpoint for team taught courses; explain to students examples of useful comments and how results are used; offer the evaluation during scheduled class time; and allow for voluntary student participation. Finally, programs should create an assessment plan that outlines when the results will be released to faculty members, who will receive the results, and what the results will be used for. Once that is decided, faculty members should be encouraged to reflect on the results, mentored on how to interpret the results, instructed on how to summarize comments, advised to longitudinally track results using tables, and required to create an action plan to address deficiencies. Ultimately, student ratings should not be the only measure of teaching effectiveness, nor conducted for summative reasons alone.³¹ Instead, multiple formative and summative measures such as self-reflection, peer evaluation, and direct student learning outcomes should also be used to evaluate and improve teaching.⁴⁶

REFERENCES

1. Anderson HM, Cain J, Bird E. Online student course evaluations: review of literature and pilot study. *Am J Pharm Educ*. 2005;69(1):Article 5.
2. Benton SL, Cashin WE. Student ratings of teaching: a summary of the literature. IDEA Paper No. 50. 2011. https://www.ideaedu.org/Portals/0/Uploads/Documents/IDEA%20Papers/IDEA%20Papers/PaperIDEA_50.pdf
3. Benton SL, Ryalls KR. Challenging misconceptions about student ratings of instruction. IDEA Paper No. 58. 2016. https://www.ideaedu.org/Portals/0/Uploads/Documents/IDEA%20Papers/IDEA%20Papers/PaperIDEA_58.pdf
4. Barnett CW, Matthews HW. Teaching evaluation practices in colleges and schools of pharmacy. *Am J Pharm Educ*. 2009;73(6):Article 103.
5. Accreditation Council of Pharmacy Education. Accreditation standards and key elements for the professional program in pharmacy leading to the doctor of pharmacy degree. <https://www.acpe-accredit.org/pdf/Standards2016FINAL.pdf>. Accessed January 9, 2019.
6. Liaison Committee on Medical Education. Functions and structure of a medical school. <http://lcme.org/publications/#Standards>. Accessed January 9, 2019.
7. Standards for accreditation of baccalaureate and graduate nursing programs. <http://www.aacnnursing.org/CCNE-Accreditation/Resource-Documents/CCNE-Standards-Professional-Nursing-Guidelines>. Accessed January 9, 2019.
8. New England Association of Schools and Colleges Commission of Institutions of Higher Education. https://cihe.neasc.org/standards-policies/standards-accreditation/standards-effective-july-1-2016#standard_five. Accessed January 9, 2019.
9. Southern Association of Colleges and Schools Commission on Colleges. The principles of accreditation: Foundations for quality enhancement. <http://www.sacscoc.org/pdf/2018PrinciplesOfAccreditation.pdf>. Accessed January 9, 2019.
10. Higher Learning Commission. Criteria for Accreditation. <https://www.hlcommission.org/Policies/criteria-and-core-components.html>. Accessed January 9, 2019.
11. Fjortoft N. A reflection on faculty and course evaluations. *Am J Pharm Educ*. 2015;79(9):Article 129.
12. Hatfield CL, Coyle EA. Factors that influence student completion of course and faculty evaluations. *Am J Pharm Educ*. 2013;77(2):Article 27.
13. Seldin P. Evaluating faculty performance: A practical guide to assessing teaching, research, and service. Boston, MA: Anker Publishing Company. 2006.
14. Schiekirka A, Raupach T. A systematic review of factors influencing student ratings in undergraduate medical education course evaluations. *BMC Med Educ*. 2015;15:30.
15. Layne L. Defining effective teaching. *J Excellence Coll Teach*. 2012;23(1):43-68.
16. DeCourcy E. Defining and measuring teaching excellence in higher education in the 21st century. *College Quarterly*. 2015;18(1):1-10.
17. Medina MS, Bouldin AS, Gonyeau M, et al. Report of the 2011-2012 Academic Affairs Standing Committee: the evolving role of scholarly teaching in teaching excellence for current and future faculty. *Am J Pharm Educ*. 2012;76(6):Article S5.
18. Glassick CE, Huber MT, Maeroff GI. Scholarship Assessed: Evaluation of the Professoriate. San Francisco, CA: Jossey-Bass, 1997.
19. Richlin L. Scholarly teaching and the scholarship of teaching. *New Directions for Teaching and Learning*. 2001;86:57-67.
20. Shulman LS. Those who understand: Knowledge growth in teaching. *Educational Researcher* 1986;15(2):4-14.
21. Glassick CE, Huber MT, Maeroff GI. Scholarship assessed: Evaluation of the professoriate. San Francisco, CA: Jossey-Bass;1997.
22. Svinicki M, McKeachie WJ. McKeachie's teaching tips: Strategies, research, and theory for college and university teachers. 14th ed. Belmont, CA: Wadworth; 2014:336.
23. Kogan JR, Shea JA. Course evaluation in medical education. *Teach Teacher Educ*. 2007;23:251-264.
24. Ryan JM, Harrison PD. The relationship between individual instructional characteristics and the overall assessment of teaching effectiveness across different contexts. *Research in Higher Education*. 1995;38:575-592.
25. IDEA. An introduction to student ratings of instruction. http://www.ideaedu.org/Portals/0/Uploads/Documents/Client%20Resources/SRI%20Infographic_Diagnostic_Form.pdf. Accessed January 9, 2019.
26. Meyer JP, Doromal JB, Wei X, et al. *Res High Educ*. 2017;58:545. <https://doi-org.ezproxy.neu.edu/10.1007/s11162-016-9437-8>.

27. Abedina NFZ, Taib JM, Jamil HMT. Comparative study on course evaluation process: Students' and lecturers' perceptions. *Procedia - Social and Behavioral Sciences*. 2014;123:380-388.
28. Schiekirka S, Reinhardt D, Heim S. Student perceptions of evaluation in undergraduate medical education: A qualitative study from one medical school. *BMC Med Educ*. 2012;12:45. <http://www.biomedcentral.com/1472-6920/12/45>
29. Cashin WE. Developing an effective faculty evaluation system. IDEA Paper No. 33. 1996 https://www.ideaedu.org/Portals/0/Uploads/Documents/IDEA%20Papers/IDEA%20Papers/Idea_Paper_33.pdf
30. Hansen WL. How a customized approach can improve teaching and learning. *Liberal Education*. 2014;100(3). <https://www.aacu.org/publications-research/periodicals/rethinking-student-course-evaluation>. Accessed January 9, 2019.
31. Benton SL, Li D. IDEA student ratings of instruction and RSVP. IDEA Paper #66, September 2017. https://www.ideaedu.org/Portals/0/Uploads/Documents/IDEA%20Papers/IDEA%20Papers/PaperIDEA_66.pdf. Accessed January 9, 2019.
32. Cashin WE. Student ratings of teaching: Recommendations for use. IDEA Paper No. 22. 1990. Center for Faculty Evaluation and Development Kansas State University.
33. Peeters MJ. Measuring rater judgment within learning assessments-part 1: why the number of categories matters in a rating scale. *Curr Pharm Teach Learn*. 2015;7:656-661.
34. Fike DS, Doyle DJ, Connelly RJ. Online vs. paper evaluations of faculty: when less is just as good. *J Effect Teach*. 2010;10(2):42-54.
35. Guder F, Malliaris M. Online and paper course evaluations. *Am J Bus Educ*. 2010;3(2):131-138.
36. Heinert S, Roberts TG. Factors motivating students to respond to online course evaluations in the College of Agricultural and Life Sciences at the University of Florida. *North Amer Colleges and Teachers of Agriculture*. 2016;60(2):189-194.
37. Fleming P, Heath O, Curran V. Making medical student course evaluations meaningful: implementation of an intensive course review protocol. *BMC Med Educ*. 2015;15(99).
38. Gimbel RW, Cruess DF, Schor K, Hooper TI, Barbour GL. Faculty performance evaluation in accredited US Public Health Graduate Schools and Programs: A national study. *Acad Med*. 2008;83(10):962-968.
39. Rutgers University Institutional Research Board website. <https://orra.rutgers.edu/rutgersstudents>. Accessed January 9, 2019.
40. Metropolitan State University Institutional Review Board website. <https://www.msudenver.edu/irb/guidance/studentsasresearchsubjects/>. Accessed January 9, 2019.
41. Smith B. Student ratings of teaching effectiveness for faculty groups based on race and gender. *Educ. Summer* 2009;129(4):615-624. *Academic Search Complete*. Ipswich, MA.
42. Mitchell KM, Martin J. Gender bias in student evaluations. *Pol Sci Politics*. 2018;51(3):648-652. <https://doi.org/10.1017/S104909651800001X>
43. Laube H, Massoni K, Sprague J, Ferber AL. The impact of gender on the evaluation of teaching: what we know and what we can do. *Nat Wom Stud Assoc J*. 2007;19(3):87-104.
44. Young S, Rush L, Shaw D. Evaluating gender bias in ratings of university instructors' teaching effectiveness. *Int J Schol Teach Learn*. 2009;3(2):Article 19.
45. Basow SA, Martin JL. Bias in student evaluations. In M. E Kite (Ed.), *Effective evaluation of teaching: A guide for faculty and administrators* (pp. 40-49). 2012. Retrieved from *the Society for the Teaching of Psychology* web site: <http://teachpsych.org/ebooks/evals2012/index.php>.
46. Hammer D, Piascik P, Medina M. Recognition of teaching excellence. *Am J Pharm Educ*. 2010;74(9):Article 164.

Appendix 1. Sample Summary Table of Students' Course Evaluation Comments

Lecturer Name: Dr. Jones	Term and Year: Fall 2017
Course Name: Clinical Communications	Course Number: PHAR XXX
Number of Students in Class: 100	Evaluation Response Rate: 60%
Theme	Count of Comments
Strengths of lecture	
Organized lecture and content	50
Measurable objectives	38
Asked questions throughout lecture; used active learning frequently	14
Handout followed slides	2
Areas of improvement	
Need to increase wait time when asking questions to class	22
Pace was fast	15
Include stories and examples to reinforce and illustrate content	1

Appendix 2. Sample Table (Template) of an Individual Faculty Member's Longitudinal Performance

Course Evaluation Results (2013-2017) (Scale: 1=low to 5=high)					
Course Name and #	Term Year	Class Size	Response Rate (%)^a	Faculty X's Mean (SD)^b	College Mean^c
Clinical Communications PHAR XXX	Fall 2017	88	50	4.5 (0.65)	4.1
	Fall 2016	83	55	4.8 (0.40)	4.2
	Fall 2015	100	60	4.3 (0.91)	4.1
	Fall 2014	109	68	4.4 (0.83)	4.1

^a % of students who completed the course evaluation

^b Faculty members' course rating on question "Overall I would rate this instructor's teaching as" (mean and SD)

^c Comparison data: Overall college faculty members' average rating on the same question